Lab 2 – MATH 240 – Computational Statistics

Charles Hooey
Affiliation
Department
`chooey@colgate.edu`

Abstract

First part of a three-part lab that involves working with the program Essentia in order to retrieve data from song libraries in order to determine how much a particular artist has contributed to a work with multiple authors. As part of task 1, I created a batch file that automatically processes the Essentia command for every song within the MUSIC directory. Following this, I became familiar with how to use the jsonlite package for R, and extracted data using this package from an example song that was provided within my assignment's repository.

Keywords: For Loops; Packages; Lists; Data Frames

1 Introduction

Week 2's lab assignment was the first part of a three-part project that involves the sourcing and extraction of music data from a program called Essentia in an effort to determine the degree of contribution among different bands collaborating on the same song. This program allows us to extract a wide variety of data from songs, though we must use its command prompt to access it. However, Essentia's command prompt must be entered for every single song within the directory that you are looking to extract data from, which can take quite some time depending on the size of the song directory. To remedy such, task one of the assignment involves creating a batch file that will automate the generation of commands for all songs which can be used to process the entire song directory at once. After creating our batch file within part one, we begin proccessing json files through the jsonlite package for R (Ooms, 2014), which allows us to extract the data for each song from Essentia.

2 Methods

As mentioned above, part one of this assignment involves the creation of a batch file in order to process every single song within our music directory through Essentia simultaneously. To create this batch file, a vector was created containing the names of all files in our music directory, so that we'd be able to reference them within R. Then, we used the strcount function from the stringr package (Wickham, 2023) in order to subset all directories within the MUSIC folder into the classifiation of artist, album, and track. Following such, two loops were established in order to iterate among all albums and song within the MUSIC folder. Within these loops, the strcount function was used to extract the name of the artist, album, and track from each song with each being assigned to an object. These objects were then pasted together with our necessary command line, and saved to an empty vector that would populate with every iteration of the loop. This code was then written to a text file, ready to be executed as a batch. For the second part of the lab, we processed an example json file using the jsonlite package for R. Using the fromJSON function, I was able to extract all Essentia data from the provided example json file. Furthermore, specific values, including: average loudness, mean of spectral energy, dancability, bpm, key, scale, and length were extracted using commands found on Essentia's website. Each type of data was then assigned to objects and printed using the paste function.

3 Results

After completing part one of this assignment, I was able to successfully create a batch file containing all songs within the MUSIC directory, which for each song included the command prompt to run Essentia, the file's name, as well as the file listed in json format. Additionally, I had also sourced and retrieved Essentia data from the example song provided in the LAB 2 repository, as well as output the song's: name, album, artist, average loudness, mean of spectral energy, dancability, bpm, key, scale, and length.

References

Ooms, J. (2014). The jsonlite package: A practical and consistent mapping between json data and r objects. *arXiv:1403.2805 [stat.CO]*.

Wickham, H. (2023). *stringr: Simple, Consistent Wrappers for Common String Operations*. R package version 1.5.1.